BY THE NUMBERS

Lorie A. Fridell

A GUIDE FOR ANALYZING RACE DATA FROM VEHICLE STOPS

Police Executive Research Forum

COPS
Community Oriented Policing Services
U.S. Department of Justice

By the Numbers:
A Guide for Analyzing Race Data from Vehicle Stops

By the Numbers:
A Guide for Analyzing Race Data from Vehicle Stops

Lorie Fridell

Police Executive Research Forum

This report was funded by the U.S. Department of Justice Office of Community Oriented Policing Services under grant 2001-CKWXK046. The points of view expressed here are those of the author and do not necessarily represent the official position of the Department of Justice, the Office of Community Oriented Policing Services or the Police Executive Research Forum or its members.

Police Executive Research Forum, Washington, D.C. 20036

Published 2004.

Printed in the United States of America

Library of Congress Number 1-878734-79-2
ISBN 2004101744

Cover art by Marnie Kenney
Interior design by Automated Graphic Systems

Contents

Foreword

The term "racial profiling" emerged only in the late 1990s, but concerns about whether some police are racially biased in their decision making date back decades, arguably even centuries, in U.S. history. The tensions between police and minority residents are longstanding and their potency is reflected by the devastating riots of the 1950s and 1960s, as well as by urban unrest in more recent years in Los Angeles, Cincinnati, Miami and myriad other locations. The good news, however, is that the law enforcement field is more capable than ever of addressing this persistent issue. We have witnessed profound reforms in policing during the last few decades—including those linked to hiring, training, policies, and community partnerships. This "new age of policing" (characterized by openness, respect for human dignity, accountability, and outreach—provides a solid foundation and a great potential for unprecedented progress in addressing the perception and practice of racially biased policing.

And, indeed, law enforcement agencies across the country have been responding to the concerns raised by their residents about racial bias. A Police Executive Research Forum (PERF) 2001 publication, supported by the U.S. Department of Justice's Office of Community Oriented Policing Services (COPS Office), was developed to help them in those efforts. That document, *Racially Biased Policing: A Principled Response*, was the first major product of the PERF/COPS partnership on this critical topic. It outlined the major areas of intervention for agencies concerned about racially biased policing and the perceptions of its practice.[1] Chapters detailed approaches for agencies in supervision and accountability, policy, recruitment and hiring,

---

1 This document can be downloaded from www.policeforum.org.

training and education, community outreach, and data collection. By choice or by mandate, many agencies have adopted this last response option (data collection. To collect data, line officers are asked to report information on vehicle and/or pedestrian stops. The information includes the race/ethnicity of the driver and other details about the stop (e.g., reasons for the stop, disposition of the stop, whether a search was conducted, outcome of the search). Data collection is meant to assist agency administrators and jurisdiction residents with determining whether racial bias influences police decisions to make stops. The chapter on data collection in the first document identified promising practices in terms of the types of stops to target for data collection and the data elements to collect. However, at that time, there had not yet emerged any promising approaches for analyzing and interpreting the data.

While the agencies that were the first to adopt data collection can be commended for their analysis efforts, PERF staff's review of early reports revealed that analyses were being conducted in a manner that did not reflect even minimal scientific standards and conclusions were being drawn that were wholly unsupported by the data. Back then, and even as this document goes to press, most agencies were and still are conducting "census benchmarking." In census benchmarking agencies compare the demographic profile of the drivers stopped by police to the demographic profile of the residents of the jurisdiction as determined by the U.S. Census. For a variety of reasons, such a comparison is of no scientific value for purposes of trying to measure racial bias in policing and, in fact, has very often resulted in misleading and unsupported findings.

Many agency executives and other stakeholders (e.g., concerned citizen leaders, civil rights leaders) have understood that the analyses of vehicle stop data is frequently wanting, but they have been frustrated by the lack of guidance in this area. Indeed, while people calling PERF and the COPS Office in the late 1990s and early 2000s inquired about the types of stops to target for data collection and the types of data to collect, by 2003

and into 2004 more calls were along the lines of, "We've collected the data; *now what do we do*?" The calls reflecting the most frustration have come from agencies that have been mandated by local or state legislation to collect data but have not been provided with the guidance or resources required to implement that charge effectively. These inquiries increased exponentially as more and more agencies voluntarily or through mandates adopted data collection. Pending state and federal legislation, if adopted, would further increase the number of agencies across the United States collecting data.

To meet the burgeoning needs of the police community, PERF and the COPS Office have partnered again, this time to provide the guidance that is needed to ensure the responsible analysis and interpretation of vehicle stop data. *By the Numbers* is a detailed "how to" guide for analyzing race data from vehicle stops. It provides a social science framework for understanding the challenges of trying to measure racial bias in policing and presents an array of methods for law enforcement professionals, researchers and other stakeholders to consider when interpreting the vehicle-stop data. The primary audience for this technical guide includes the people who will actually be conducting the analyses, though police professionals at all levels, policy makers and others have much to gain by reading the preliminary chapters. Following these introductory chapters, *By the Numbers* provides step-by-step guidance for implementing various benchmarking methods. A companion document, *Understanding Vehicle-Stop Race Data: A Stakeholders' Guide*, is specifically written for a broader audience—including police agency executives, concerned residents, advocacy groups, the media, and policy makers. It discusses the challenge of analyzing vehicle-stop data and summarizes the key contents of By the Numbers in a less technical fashion.[2]

---

2 This guide can be downloaded from www.policeforum.org in 2005.

By the Numbers does not recommend a "perfect method" that will allow an agency to simply, easily and definitively measure whether racial bias is manifested in police decisions. Such a method does not exist. The question of whether bias influences some officers when they stop drivers, like many other social science research questions in criminal justice and related fields, is impossible to answer with complete certainty. There are, however, some methodologies that are much stronger than others in their ability to answer the key research question—that is, we can have more confidence in the results. *By the Numbers* not only provides detailed guidance for implementing the various methods, but also includes assessments of each method's strengths and weaknesses. This information will help readers implement the strongest method that available resources will allow and ensure that the conclusions they draw from the chosen method are responsible.

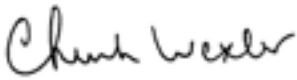
While this technical document is meant to assist the individuals who are conducting the analyses, it is designed to serve many others. PERF and the COPS Office hope that this document will ensure that residents and other stakeholders receive responsible answers to their very real questions about racial bias. In addition, we hope that this document will assist policy makers who will make decisions regarding whether to mandate that agencies in their jurisdiction collect data. These men and women can now make this decision with an understanding of the challenges of measuring racial bias and the considerable resources required for the responsible implementation of a data collection mandate.

PERF and the COPS Office hope that this document will be of value to law enforcement practitioners. The vast majority of police officers in this country are principled, dedicated men and women who are committed to serving all citizens with equity and fairness. They now find themselves the "subjects" of study by virtue of voluntary or mandatory data collection. It is unjust to have their reputations tarnished by non-scientific analyses and we trust this document will help prevent this

practice and, instead, ensure responsibly implemented and reported analyses of vehicle-stop data.

Beyond bringing value to individual audiences of concerned residents and law enforcement practitioners, we hope this document will facilitate stronger relationships between them. The issues related to racially biased policing and the perceptions of its practice cannot be addressed effectively by either group alone. To address this longstanding issue, residents, other stakeholders and police must join together to identify concerns about law enforcement practices and outline how they will be resolved. PERF and the COPS Office hope this document will substantially advance this important dialogue.

Carl Peed
Director, COPS Office



Chuck Wexler
Executive Director, PERF

Acknowledgements

On behalf of PERF, I convey great appreciation to Director Carl Peed, Deputy Director Pam Cammarata, and other leaders in the U.S. Department of Justice Office of Community Oriented Policing Services (COPS Office). They shared our belief that there was a significant need for this document and provided the necessary funding to bring it to fruition. I am particularly grateful for the very constructive guidance and unwavering support I received throughout this project from Tamara Clark Lucas and Amy Schapiro, our COPS Office monitors.

An advisory board made up of both practitioners and social scientists met early on in the project to guide the content and form of this document. Many read chapters or sections as they were completed. Members of this board included Dr. Gary Cordner, Eastern Kentucky University; Dr. Scott Decker, University of Missouri, St. Louis; Assistant Chief John Diaz, Seattle (WA) Police Department; Captain Thomas Didone, Montgomery County (MD) Police Department; Dr. Don Faggiani, PERF; Dr. Amy Farrell, Northeastern University; Secretary Ed Flynn, Massachusetts Executive Office of Public Safety; Assistant Chief John Gallagher, Miami (FL) Police Department; Former Chief Mark Kroeker, Portland (OR) Police Bureau; Dr. John Lamberth, Lamberth Consulting; Dr. Catherine Lawson, State University of New York at Albany; Dr. Steve Mastrofski, George Mason University; Dr. Jack McDevitt, Northeastern University; Antony Pate, COSMOS Corporation; Assistant Deputy Superintendent Karen Rowan, Chicago (IL) Police Department; Professor Margo Schlanger, Harvard University; Dr. Ellen Scrivner, Former Deputy Director of the COPS Office; Dr. Deborah Thomas, University of Colorado at Denver; Dr. Sam Walker, University of Nebraska at Omaha; Dr. Charles Wellford, University of Maryland; and Dr. Matt Zingraff, North Carolina State University.

A group of academics—most of them members of the advisory board—stand out for their contributions to the content of this report. They are the key social scientists around the country who are analyzing and interpreting police-citizen contact data for various law enforcement entities and who have been instrumental in advancing the methods used to assess racial bias. Members of this group were generous with their time and knowledge. They shared their draft and completed reports from their own research, responded to phone and email inquiries, served as a "sounding board" for particularly vexing issues, and reviewed critically the various chapters of this document that reported on their work and/or otherwise reflected most closely their experience and expertise. In alphabetical order, these people are

Geoff Alpert, *University of South Carolina*
Gary Cordner, *Eastern Kentucky University*
Scott Decker, *University of Missouri at St. Louis*
Robin Engel, *University of Cincinnati*
Amy Farrell, *Northeastern University*
David Harris, *University of Toledo*
John Lamberth, *Lamberth Consulting*
Catherine Lawson, *State University of New York at Albany*
Jack McDevitt, *Northeastern University*
Michael Smith, *Washington State University*
Samuel Walker, *University of Nebraska at Omaha*
Matthew Zingraff, *North Carolina State University*

It is no exaggeration to say that there would be no content for this document without their research efforts in the field and no summary of the promising approaches without their direct assistance in writing this document.

Appendix A was written by Karen Parker of the University of Florida and Appendix D was contributed by John Lamberth, David Harris, Jack McDevitt and Deborah Ramirez. Providing additional expert review of particular chapters or sections were Ian Ayres of Yale University; R. Richard Banks of Stanford Law School; Deputy Chief Michael Berkow of the Los Angeles Police

Department; Jennifer Calnon of Pennsylvania State University; Jerry Clayton and Karl Lamberth of Lamberth Consulting; Assistant Director Rod Covey of the Arizona Department of Public Safety; Scott Henson of the ACLU of Texas; Chief Ellen Hanson of the Lenexa Police Department; Chief Glenn Ladd of the North Kansas City Police Department; Nicholas Lovrich of Washington State University; Mike Maltz of the University of Illinois at Chicago; Ken Novak of the University of Missouri, Kansas City; Tim Oettmeier, Executive Assistant Chief of the Houston Police Department; Nicola Persico of the University of Pennsylvania; Jeff Rojek of St. Louis University; Matt Salazar of the Austin Police Department; and Jane Wiseman of the Massachusetts Department of Public Safety. Jerry Deichert at the University of Nebraska at Omaha provided expert advice on census data. To all of these people, I convey my sincere appreciation. And my thanks go as well to two "mystery reviewers" who provided very detailed and constructive feedback on many chapters but who prefer to remain anonymous. They know who they are.

I owe a tremendous debt to Barbara deBoinville who provided expert editing—ensuring that complicated, sometimes dense material was organized, accessible and clear. At PERF my thanks go to Executive Director Chuck Wexler who pushed me to complete the document while being completely supportive with the encouragement and resources I needed to do so. He also provided important insights related to the content. I'm grateful to Martha Plotkin and David Edelson who provided their usual high-quality, expert advice on the document's form and content and oversaw all aspects of Web and print publication. A colleague returning to PERF, Don Faggiani, reviewed various sections of the report and was on call day and night to provide advanced methodological guidance. Finally, special thanks to all of my staff in the Research Unit at PERF who are wonderful in many, many ways and, for this particular project, thoughtfully and patiently let me hole up without interruption for days at a time to write. This one's for you.

Lorie A. Fridell

Lorie Fridell

Introduction

Law enforcement agencies across the country are attempting to address the issues of racially biased policing and the perceptions of its practice. Racially biased policing is here defined as the inappropriate consideration by law enforcement of race or ethnicity[1] in deciding with whom and how to intervene in an enforcement capacity.[2] Decades of profound reform reflected in community policing are threatened by perceptions of racially biased policing and its practice. This trust-shattering issue is placing at risk the partnerships with residents, particularly minority residents, that police have worked diligently to develop. At the same time, however, it is these very partnerships—if they are solid—that can provide the basis for effective reforms. In short, these partnerships with the community provide law enforcement agencies with the general capabilities and specific tools they need to address these critical issues.

In 2001 the Police Executive Research Forum (PERF), with funding from the U.S. Department of Justice, Office of

---

1 Mirroring the U.S. Census we use "ethnicity" in this document to refer to whether a person is of Hispanic or non-Hispanic origin.

2 For a discussion of the various policy options that, in effect, define "inappropriate," see the web site of the Police Executive Research Forum, www.policeforum.org.

Community Oriented Policing Services, published *Racially Biased Policing: A Principled Response*. This report (Fridell et al. 2001) outlined the various ways that law enforcement agencies can effectively address racially biased policing. Specifically, it discussed methods of reform and prevention in the areas of accountability and supervision, policies to address racially biased policing, recruitment and hiring, education and training, minority community outreach, and collection of data on police-citizen contacts.[3]

As part of their comprehensive response to the issues related to racially biased policing, many law enforcement agencies are collecting data on various types of police-citizen interactions, including information regarding the race and ethnicity of persons stopped by police. PERF's first report on racially biased policing (Fridell et al. 2001) discussed the pros and cons of data collection and provided guidance to agencies mandated or choosing to collect data regarding the types of activities to target (for example, traffic stops, investigative stops) and the specific data to collect (for example, date/time of stop, reasons for stop). The topic of this new report is the analysis and interpretation of the vehicle stop data collected by agency personnel.[4] One purpose is to describe the social science challenges associated with data collection initiatives so that agencies and other stakeholders can be made fully aware of both the potential and limitations of police-citizen contact data. The second purpose is to provide a "how to" guide for the analysis/interpretation of the data so that the jurisdictions that are collecting it can con-

---

3 This report, funded by COPS Office Grant 1999-CK-WX-0076, is available in its entirety on the PERF web site at www.policeforum.org.

4 The methods we describe pertain to the analysis of vehicle stops (not pedestrian stops) of all types. The term "vehicle stop" is used to denote any stop made by police of a person in a vehicle. The term "traffic stop" denotes a vehicle stop the stated purpose of which is to respond to a violation of traffic laws (including codes related to quality/maintenance of vehicles). The term "investigative (vehicle) stop" denotes police stops of people in vehicles when there is reasonable suspicion of criminal activity.

duct the most valid and responsible analyses possible with the resources they have. This report will be of greatest value to the people charged with analyzing the data. They include law enforcement agency research staff, outside social scientists, interest group members, or other stakeholders.[5]

A companion document entitled *Understanding Race Data from Vehicle Stops: A Stakeholder's Guide* (Fridell forthcoming, 2005) is geared more broadly for police practitioners; concerned residents; advocacy groups; the media; and local, state, and federal policy makers. Its purpose is to educate this wide audience about the potential and constraints associated with data collection efforts. It discusses the challenge of benchmarking, how to assess the quality of benchmarks, how to interpret results responsibly, and how to use the data for constructive dialogue and reform.[6]

Law enforcement agencies' documents reporting on the results of their data analysis efforts provided an important source of information for this report. In these documents PERF staff identified promising procedures and methodologies as well as common weaknesses and missteps.[7] Additionally, PERF staff relied upon the valuable expertise of an advisory board. Its

---

5 For purposes of simplification, throughout this document we refer to the "agencies" or "agency researchers" conducting analyses, although we acknowledge that researchers outside or independent of the agency may be analyzing jurisdiction data.

6 Many of the topics treated in the companion document are also covered in Chapter 2 of this report. A related resource, funded by the Office of Community Oriented Policing Services, is *How to Correctly Collect and Analyze Racial Profiling Data: Your Reputation Depends on It!* (McMahon et al. 2002). This document is available through the COPS Office web site at www.cops.usdoj.gov.

7 The documents (for example, jurisdictions' reports of their results) that manifested weaknesses or missteps are not mentioned by name in this report. (The reader will always find references to strong studies and documents.) We saw no constructive purpose in publicly linking faulty work to specific agencies or researchers, most of whom generously provided PERF with their materials for review.

members—listed in the acknowledgments section—include the key social scientists around the country who are analyzing and interpreting police-citizen contact data, experienced law enforcement practitioners, and personnel within research units of law enforcement agencies. Members of this advisory board provided PERF with the documents they had written or commissioned on the methods and results of data analysis/interpretation. Board members helped define the contents of this report and reviewed early drafts. Therefore, the pronoun "we" is used throughout the report to acknowledge that its contents reflect this collective wisdom.

Chapter 2 describes the social science challenges associated with analyzing and interpreting the police-citizen contact data collected to measure racially biased policing; specifically, it explores the goal, the potential, and the limitations of what has come to be called "benchmarking" the data. Chapter 3, "Getting Started," explains the steps agencies should take when they initiate collection and analysis of police-citizen contact data, including how to develop a data collection plan, how and why to involve residents and police personnel from all levels of the agency, and how to select benchmarks. Chapter 4 examines issues that are relevant to all analysis efforts, regardless of their particular focus or the benchmarking method selected. Topics include reviewing data quality, selecting reference periods, and analyzing subsets of data.

Chapters 5 through 10 present information on methods that can be used to address the first of two research questions:

- Does a driver's race/ethnicity have an impact on vehicle stopping behavior by police?

In considering this question, a researcher is attempting to assess whether racially biased policing is manifested in the decisions of officers regarding whom to stop. In Chapter 11, we address the second research question:

- Does a driver's race/ethnicity have an impact on police behaviors/activities during the stop?

With regard to this research question, we describe how to assess the impact of race/ethnicity on the activities that occur after the stop is made. Most importantly, we discuss how to examine the disposition of the stop and search activity.

In Chapter 12 we suggest to the readers who are not advanced statisticians what calculations to use to measure disparity between racial/ethnic groups. In Chapter 13 we discuss how to use the results from data collection to achieve reform.

Chapters 1 through 3 present important information for all people who are stakeholders in the collection of police-citizen contact data; Chapter 13, on using the data for reform, is also geared toward this broad audience.[8] The material in between fills the need—identified by the Office of Community Oriented Policing Services and PERF—for very specific and technical information regarding how to analyze and interpret these data.

"Best practices" in analyzing and interpreting police-citizen contact data are continuing to evolve as social scientists make progress in this area. Because of these advances, PERF will retain a web site that will provide new resources and information as they become available.[9]

---

8 The companion document, which is geared to a wide variety of stakeholders, conveys much of this material as well. It also summarizes material in Chapters 4 through 12 in a less technical fashion.

9 See www.policeforum.org.

2 The Benchmarking Challenge

Jurisdictions collecting police-citizen contact data are calling upon social science to determine whether there is a cause-and-effect relationship between a driver's race/ethnicity and vehicle stopping behavior by police. In analyzing the data, researchers have attempted to develop comparison groups to produce a "benchmark" against which to measure their stop data. If an agency determines that, say, 25 percent of its vehicle stops are of racial/ethnic minorities, to what should this be compared? In other words, what percentage would indicate racially biased policing?" This is the question at the core of benchmarking. To determine an answer, researchers have compared the demographic profiles of people stopped by police to the demographic profiles of the residential population of the jurisdiction, to the demographic profiles of residents with a driver's license, and to the demographic profiles of people observed driving on jurisdiction roads—to name a few comparison groups.

THE OBJECTIVE OF BENCHMARKING

Before we discuss the various methods for benchmarking, it is constructive to consider our objectives when analyzing police-citizen contact data. Then we can outline how benchmarks vary in their ability to achieve these objectives. We start

with two conceptual models. Figure 2.1 shows a model of the first research question: Does a driver's race/ethnicity have an impact on the decisions police make with regard to whom to stop? We want to know if X (driver race/ethnicity) has any causal impact on Y (police decisions to stop drivers). To determine causality, however, we must exclude or "control for" rival causal factors—factors other than the race/ethnicity of the driver—that could explain police stopping decisions (see the model in Figure 2.2). In attempting to test whether X causes Y, we need to rule out alternative hypotheses that A, B, C, and Z—either alone or together or in interaction with X—cause Y.

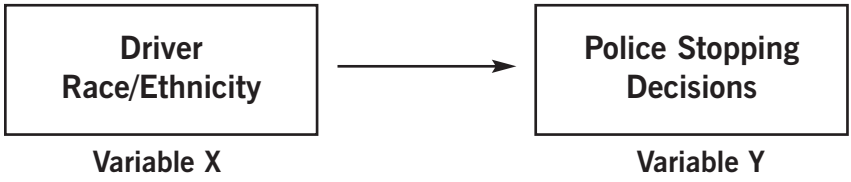


Figure 2.1. Model of First Research Question: Does Driver Race/Ethnicity Affect Vehicle Stopping Decisions Made by Police?

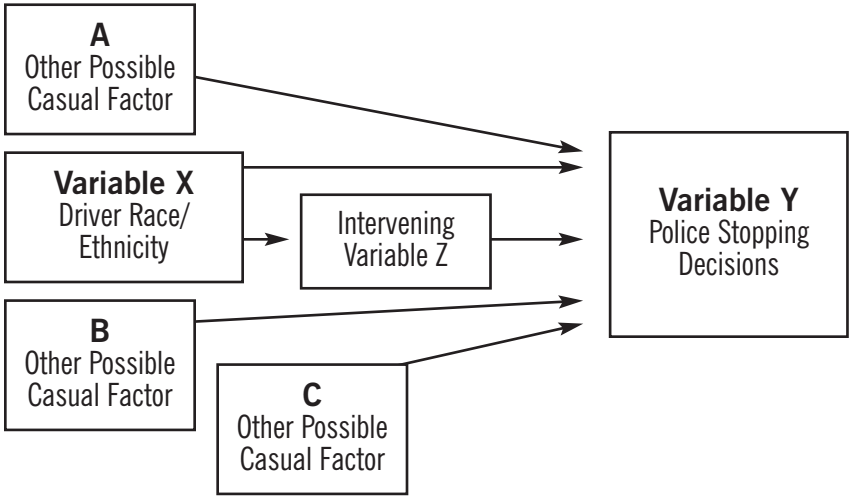


Figure 2.2. Model of Factors, Other than Bias, that Might Affect Stopping Decisions Made by Police

The following example clarifies why rival causal factors must be ruled out in any analysis of police-citizen contact data. Let us say that parents are concerned that the grading by math teachers at a high school reflects teachers' bias against females. The parents' allegation is that these math teachers believe boys are better than girls at math and that—consciously or unconsciously—these attitudes are reflected in the grades being given to the students.

Our basic conceptual model is that gender (X) has a causal impact on grades (Y). To test this scientifically, however, we cannot conduct analyses that consider only X and Y. We cannot, for instance, look only at the percent of females who got A's and B's and the percent of males who got A's and B's and draw any conclusions regarding teachers' gender bias. Instead, we must consider other factors that affect grading behavior. A key variable, of course, would be students' math performance. Our analyses must control for math performance (for example, scores on objective tests). In other words, our research design or statistical techniques must remove or "neutralize" the impact of performance on grades. If, after we have controlled for math performance, we still find that males get better math grades than do females, then we must seriously consider the possibility of gender bias by teachers.

Now let us return to the first research question concerning who is stopped by police. Police can have various legitimate reasons for deciding to stop a vehicle. These reasons are the rival causal factors that would become the A, B, and C of Figure 2.2. Let's again consider gender but in the context of analyzing police stopping behavior, not math grades.

The reports of most jurisdictions regarding their police-citizen contact data state that males are stopped by police more than females. For instance, a jurisdiction may find that 65 percent of its vehicle stops by police are of male drivers and 35 percent are of female drivers. Does this indicate gender bias on the part of the police? It is unclear from these data, but most of us are disinclined to jump to that conclusion because we can think

of factors other than police bias that could account for the disproportionate stopping of male drivers. That is, alternative hypotheses for the data exist. One possibility is that men drive more than women (the quantity factor). Another possibility is that men violate traffic laws more often than women do (the quality factor). A third possibility is that more males than females drive in the areas where police stopping activity tends to occur (the location factor). We do not know if these possibilities are true, but we must consider these alternative explanations in our research design because it is logical to assume that

- people who drive more should be more at risk of being stopped by police,
- people who drive poorly should be more at risk of being stopped by police, and
- people who drive in locations where stopping activity by police is high should be more at risk of being stopped by police.

For the purposes of our example, the objective of benchmarking is to see if gender bias is at work. If we could develop a gender profile of the people who *should be* more at risk of being stopped by police, we could compare it to the gender profile of the people who *are* being stopped by police. That is, if we managed through our research design to determine that men *should comprise* 65 percent of the police stops because of their driving quantity, quality, and location, and if indeed they *do comprise* 65 percent of the police stops (based on the stop data collected), then we could conclude that gender bias was not affecting stopping behavior by police.

Benchmarking is the essential tool used by researchers in their quest to develop a racial/ethnic profile of the people who should be at risk of being stopped by police, assuming no bias. The variation in quality across benchmarks is directly related to how closely each benchmark represents the group of people who should be at risk of being stopped by police if no bias

exists. The following example will help clarify what we mean by benchmark quality. If a researcher uses road-side observers to develop a demographic profile of drivers who violate traffic laws, the researcher has produced a benchmark that represents fairly well the group of people who should be at risk of being stopped by police if no bias exists. On the other hand, if that same researcher used instead U.S. Decennial Census data to develop a demographic profile of people who live in the jurisdiction, the researcher has produced a benchmark that does not represent well the people at risk of being stopped by police if no bias exists. The next section on the bias hypothesis and the alternative hypotheses expands upon this discussion of benchmark quality. As we will demonstrate in this report, the variation in quality across benchmarks is great.

THE BIAS HYPOTHESIS AND THE ALTERNATIVE HYPOTHESES

Here we introduce the alternative hypotheses (hypotheses other than the one that reflects the possibility of police bias). Law enforcement agencies should consider these hypotheses when analyzing the police-citizen contact data they have been mandated to collect or have voluntarily collected to measure whether racially biased policing exists in their jurisdiction. The hypotheses reflect drivers' driving quantity, quality, and location—the factors that could legitimately influence whom police stop. This list of hypotheses will become a tool in the chapters ahead for evaluating each benchmarking method. We will indicate which of the alternative hypotheses are adequately addressed in each benchmark.

Again we want to know what the demographic profile of drivers stopped by police would look like assuming no bias. Starting at the very basics to make our point, we might ask why—in a jurisdiction made up of Caucasians, African Americans, Hispanics, and Asians—the police do not report that 25 percent of their traffic stops are of Caucasians, 25 percent are of African Americans, 25 percent are of Hispanics, and

25 percent are of Asians? One hypothesis is that *police are racially/ethnically biased in their decisions regarding whom to stop*. Competing alternative hypotheses are as follows:

- racial/ethnic groups are not equally represented as residents in the jurisdiction
- racial/ethnic groups are not equally represented as drivers on jurisdiction roads
- racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior
- racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.

In order to draw valid conclusions regarding whether racial bias is occurring, we would need to rule out all other possible, legitimate explanations for disparity. Ideally, our analysis and interpretation of stop data would encompass all of the factors reflected in those alternative hypotheses.

If we address the second hypothesis—*racial/ethnic groups are not equally represented as drivers on jurisdiction roads*—we need not concern ourselves with the first hypothesis—racial/*ethnic groups are not equally represented as residents in the jurisdiction*. That is, for purposes of identifying who is at risk of being stopped by police in a vehicle, if we know who is driving on jurisdiction roads, we do not need to know who lives in that jurisdiction. Similarly, addressing the third hypothesis—*racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*—arguably negates the need to address the first two. It can be argued that knowing who is engaging in law-violating behavior negates the need to know who is on the road. Police are not told to pull over "people on the road" but rather "people who are violating laws." The fourth hypothesis—*racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high*—stands alone and must be addressed independently of the other three. Each will be discussed below.

Research has shed light on the alternative hypotheses. This information is important because it shows us that we cannot ignore these hypotheses and presume no differences exist between racial/ethnic groups. (That is, we cannot presume the null hypothesis.) For each of the hypotheses, there is evidence that differences do exist between groups, or at least there is insufficient information to prove to any acceptable degree of certainty that no differences exist. Unless research shows there are no differences between groups as pertains to these hypotheses, we must assume that there are differences. Again this requires researchers to use methods that consider the factors encompassed in the alternative hypotheses or, at the very least, interpret their results responsibly in light of any deficiencies in their chosen methodology.

Hypothesis 1: Racial/ethnic groups are not equally represented as residents in the jurisdiction.

The demographic profile of people who live in a jurisdiction will affect the demographic profile of the people who are driving on the jurisdiction's roads. Thus, the above hypothesis is indirectly related to the "quantity" factor, and we need to include it in anticipation of our later discussion of census benchmarking (a comparison of the demographic profile of people stopped by police to the demographic profile of jurisdiction residents as measured by the U.S. Census Bureau). That racial/ethnic groups are not equally represented among residents in jurisdictions is, of course, quite obvious to all. According to the 2000 Decennial Census, 75.1 percent of the U.S. population is White, 12.3 percent is Black or African American,[1] and 3.6 percent is Asian; 9.0 percent of the population self-identify as being of more than one race. Just over 12 percent (12.5 percent) of U.S. residents (of all races) are of Hispanic origin. Although figures for different jurisdictions will deviate from this breakdown of the total U.S. population, we can

---

1 African American and Black are used interchangeably for the purposes of this document.

confidently state that no jurisdiction has equal representation in its population of racial/ethnic groups.

Hypothesis 2: Racial/ethnic groups are not equally represented as drivers on jurisdiction roads.

Not only are racial/ethnic groups not equally represented among residents in the jurisdiction (the alternative hypothesis mentioned first), but their representation as residents might not match their representation as drivers using jurisdiction roads. This might be because of (1) racial/ethnic differences in driving quantity and/or (2) racial/ethnic differences in the population of people who do not live in the jurisdiction but drive in it. This is relevant to the analysis of vehicle stops by police. If one demographic group has more presence on the road than another, it should be more at risk of being stopped.

Driving Quantity

There is evidence that racial/ethnic groups differ in the amount of their driving. National data from the U.S. Decennial Census and from the National Household Transportation Survey (NHTS) indicate that racial/ethnic minorities are under-represented as drivers relative to their residential populations.[2] The U.S. Decennial Census provides data on the percent of households that do not own vehicles, an indirect measure of driving quantity. In his comprehensive report on commuting patterns based on 1990 Census data, Pisarski (1996, xv) reports that "on average, more than 30 percent of Black households do not own vehicles, and in central cities the number is over 37 percent."[3]

---

2 The National Household Transportation Survey (previously called the Nationwide Personal Transportation Survey and the American Travel Survey) is conducted by the U.S. Department of Transportation. See www.bts.gov/nhts.

3 Some cities have "extraordinary levels of Black households without vehicles" (Pisarski 1996, 36). In New York, 61 percent of Black households are without vehicles. The corresponding figures for Philadelphia, Chicago, and Washington, D.C., are 47 percent, 43 percent, and 43 percent, respectively.

Nationally, 19 percent of Hispanic households do not own vehicles; in central cities that number rises to 27 percent. In contrast, just under 9 percent of White non-Hispanic households are without vehicles, with a corresponding figure of 15 percent for central cities (Pisarski 1996, 36).

Vehicle ownership is an indirect measure of driving quantity. Information from the National Household Transportation Survey provides more direct measures of driving quantity. Its data indicate that nonminorities drive more than minorities. For instance, the 1995 NHTS indicated that African Americans average fewer "trips per day" (including fewer vehicle trips) than do Caucasians and that Hispanics are twice as likely as non-Hispanics to use public transportation (instead of privately owned vehicles).

While the 2000 Census data on vehicle ownership and NHTS data on driving quantity both imply that minorities are under-represented as drivers relative to their representation in the U.S. population, other research reminds us that this is not going to be true in all places at all times. For instance, research conducted by the United Kingdom's Home Office (MVA and Miller 2000) found that minorities were over-represented as drivers relative to their representation in the residential populations in the areas studied.[4] In Sacramento, California, Howard Greenwald compared the demographic profiles of drivers at various intersections (using observation) to the demographic profiles of residents in the same areas (using census data); he found over-representation of minorities as drivers in some areas and under-representation of minority drivers in others (Greenwald 2001). These two small-scale studies, although of less weight than the large-scale research findings of the NHTS and U.S. Census, nonetheless support our simple point: jurisdiction-level studies of racially biased policing must consider

---

4 The Home Office of the United Kingdom is the government department responsible for promoting safe communities. Its closest equivalent in the United States is the National Institute of Justice.

the possibility that racial/ethnic groups are not equally represented as drivers on jurisdiction roads because of differences in their quantity of driving.

The need to consider the extent to which the various racial/ethnic groups are driving on the roads becomes more clear in the context of a recommendation that we will make repeatedly throughout this report—namely that researchers should conduct analyses for geographic "subareas" of the jurisdictions they are studying. Researchers are cautioned not to conduct a single analysis for the entire jurisdiction but numerous analyses within the various subareas. Within this context, it becomes more obvious why researchers should consider the extent to which each racial/ethnic group is driving on the particular roads of a subarea. Whereas it may be true (as the various large-scale studies described above indicate) that for the jurisdiction as a whole, minority representation on the roads is less than for Caucasians, this certainly will not be true for all subareas. Indeed, in some areas, minorities will be the predominant group on the roads.

Driving by Nonresidents

There is another reason—other than differences in driving quantity of jurisdiction residents—that racial/ethnic groups may not be equally represented as drivers on jurisdiction roads (and why their representation on the roads may not reflect their representation as residents). Racial/ethnic groups may not be equally represented among the *nonresidents* who drive in the jurisdiction; that is, racial/ethnic groups may not be equally represented among the people who live outside of the jurisdiction but drive into it.[5] The extent to which nonresidents drive with-

---

5 In its first annual report regarding police-citizen contact data, the Denver Police Department (Thomas 2002) revealed that 62.5 percent of the Whites stopped in their vehicles by police were nonresidents compared to 32.8 percent of the Blacks who were stopped and 35.2 percent of the Hispanics who were stopped.

in the jurisdictions that are collecting police-citizen contact data will vary greatly, as might the demographic profile of those drivers. The influx of nonresident drivers will be particularly significant in the big cities that draw commuters in from surrounding jurisdictions, especially the suburbs, during the daytime hours.[6] Additionally, nonresidents will drive into the "target jurisdiction" (the jurisdiction that is the subject of police-citizen contact data analysis) to shop, seek entertainment, vacation, travel on to another jurisdiction, and for other reasons. These nonresident drivers will affect the demographic profile of drivers on the roads of the target jurisdiction.

Clearly, the hypothesis that *racial/ethnic groups are not equally represented as drivers on jurisdiction roads* is a viable alternative hypothesis that should be accounted for in the analysis of police-citizen contact data. This report will describe how law enforcement agencies can incorporate this alternative hypothesis into their study design.

Hypothesis 3: Racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.

Driving behavior is a critical component of any model that seeks to explain decisions by police to stop drivers. Indeed, police are asked to make driving behavior a key part of these decisions, and therefore we must recognize this variable in our methodology unless we are quite confident that there are no differences across racial/ethnic groups. Excluding driving behavior from the model is equivalent to excluding math performance from the earlier analysis that tested gender bias in math teachers.

---

6 In 1993, 43 percent of the traffic tickets given in Seattle were given to nonresidents (Scales 2001). The Denver Police Department (Thomas 2001) reported that from June 2001 through May 2002 (the reference period for its second summary report) over one-half of its traffic stops were of nonresidents. In Louisville (Edwards et al. 2002a) and Iowa City (Edwards et al. 2002b), fewer than two-thirds of all drivers stopped were city residents.

It is possible, according to this hypothesis, that vehicle stopping behavior by police may not be equivalent across racial/ethnic groups because racial/ethnic groups violate traffic laws at different rates or at different levels of seriousness. These possibilities must be recognized. Concerned stakeholders have questioned the inclusion in our analysis of the third hypothesis (*racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*). They have asked the author whether the unstated implication is that minorities violate more. Indeed, no direction is implied by its inclusion. Minorities may violate traffic laws with less frequency than do majority populations. (In fact, this could be the case in light of minorities' concern about racial profiling and the increased attention they perceive they get from police.) If minorities do violate less, then it is important that this information be incorporated into the analysis to appropriately determine the rate at which they should be stopped by police in light of their driving quality. Driving behavior cannot be removed from our analysis unless there is clear evidence in support of the null hypothesis (no differences between racial/ethnic groups exist). The following information calls the null hypothesis into question.

Information on the Equivalence of Driving Behavior

The scarcity of large-scale quality research on driving behavior and race/ethnicity does not negate the importance and viability of this alternative hypothesis. In fact, it does just the opposite: what is important for our purposes is the absence of sufficient research to *rule out* the possibility of racial/ethnic differences in the nature and extent of law-violating behavior. Again, even if we had national data pointing to equivalent driving behavior or pointing to one particular direction or the other, we could not presume that those results were applicable to all times and all places.

The information on the equivalence of driving behavior across racial/ethnic groups is limited and mixed. There is research in the transportation field, albeit not substantial, indicating some differences across racial groups with regard to certain traffic violations. For instance, Feest (1968) found that Whites

were more likely than minorities not to stop at stop signs. Other researchers analyzing police-citizen contact data have produced information indicating other differences in violating behavior across racial/ethnic groups. For instance, Lange, Blackman, and Johnson (2001) found that along segments of the New Jersey turnpike where the speed limit was 65 miles per hour rather than 55 miles per hour, African Americans were disproportionately represented among the few speeders.[7] In contrast, Lamberth (1996a, 1996b) conducted research in New Jersey and Maryland and found no differences in the demographics of speeders versus non-speeders. He reports that all racial/ethnic groups were speeding in high, and similar, proportions.[8]

In citing these mixed findings, we are not trying to argue that there are differences in violating behavior across racial/ethnic groups. Quite the contrary: we do not know whether differences exist or not. Because the research does not allow us to rule out the possibility of differences in driving quality across racial/ethnic groups, we contend that research analyzing police-citizen contact data should address the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*.[9]

Youthfulness and Driving Behavior

Youthfulness has been linked to law-violating behavior. If a racial/ethnic group has proportionately more young people than

---

7 This study was criticized for various aspects of its methodology and the high proportion of missing data produced by those methods.

8 These studies defined speeding so broadly (1 mile per hour over the speed limit in Maryland and 5 miles per hour over the speed limit in New Jersey) that speeders included most drivers. This broad definition reduced the researcher's ability to detect any existing, finer distinctions in driving behavior across groups.

9 A challenge to this view is presented in Appendix D in the context of discussing the observation method of benchmarking.

another, age becomes an important "intervening variable"[10] in the analysis model. (It is a potential "Variable Z" in Figure 2.2.) We must consider whether the breakdown of age groups in a jurisdiction (or in the subareas being analyzed) varies across racial/ethnic groups. For example, if 30 percent of the minority population of an area is young (24 years of age or less) and only 20 percent of the Caucasian population is young, this phenomenon would lead to more drivers who violate the law in the minority population than in the nonminority population, assuming the link between poor driving and age.

An example using (extreme) hypothetical data will convey the potential impact of this circumstance (unequal proportions of young people within racial/ethnic groups) on police-citizen contact data being analyzed to measure racially biased policing. Table 2.1 shows the representation of Caucasian and minority drivers on the road and among those stopped by police in hypothetical Jurisdiction Q. There were 1,000 Caucasian drivers and 1,000 minority drivers on the road during the data collection period. That is, Caucasians and minorities each made up 50 percent of the driving population. Among the Caucasian drivers, 300 or 30 percent were between the ages of 15 and 24, and 700 or 70 percent were 25 or older. (We use age 15 as the lower cut-off point to include only people of driving age.) The corresponding percentages for the minority group of drivers were 60 percent and 40 percent. That is, 600 of the drivers were between the ages of 15 and 24, and 400 were 25 years of age or older.

The police in hypothetical Jurisdiction Q are completely devoid of racial/ethnic bias, and they legitimately stop, as a result of the drivers' poorer quality driving, two times as many drivers between the ages of 15 and 24 as drivers 25 years of age and older. (To make our point, we assume equivalence of driving behavior across racial/ethnic groups.) Twenty percent of the young

---

10 We use the term "intervening variable" to refer to a variable (measured or unmeasured) that is linked causally to one or more other variables in an equation or model.

Table 2.1. Representation of Caucasian and Minority Drivers in the Driving Population and Population of Stopped Drivers, by Age, Hypothetical Jurisdiction Q

Age Group	Caucasians (n=1,000)		
	Number of Drivers	Percent Stopped	Number Stopped
15-24	300	20%	60
25+	700	10%	70
Total	1,000	13%	130
Percentage of all stops: 45.61%			
Age Group	**Minorities (n=1,000)**		
	Number of Drivers	Percent Stopped	Number Stopped
15-24	600	20%	120
25+	400	10%	40
Total	1,000	16%	160
Percentage of all stops: 56.14%			

Caucasians were stopped (0.2 x 300 = 60), and 20 percent of the young minorities were stopped (0.2 x 600 = 120). They stopped 10 percent of the Caucasian drivers age 25 or above (producing 70 stops) and 10 percent of the minority drivers age 25 or above (producing 40 stops). The effect of the differential representation of young people among the minority drivers can be seen when we look at the overall representation of Caucasians and minorities among the drivers stopped by police (Figure 2.3). Caucasians made up 50 percent of the drivers (1,000 of the total 2,000) and only 46 percent of the stops. Minorities made up the other 50 percent of the drivers but 56 percent of the stops. Even though racial bias is not manifested by the police (equivalent stopping behavior across racial/ethnic groups), our data indicate (falsely) that disparity exists. If the researcher for Jurisdiction Q did not, as we did, analyze the data within age groups to confirm a lack of disparity, the researcher would have mistakenly concluded that there was disparity across racial groups. The disproportionate

representation of youth in the minority population and the increased likelihood of young people being stopped by police produced the misleading results shown in Figure 2.3: minorities appeared to be over-represented among people stopped relative to minorities' representation in the driving population.

In sum, the strongest research methodologies will address the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.* Theoretically, driving behavior is quite relevant to decisions by police to stop drivers, and the research that has been conducted on the relationship between driving quality and race/ethnicity is not sufficient for us to assume no differences across groups. Complicating matters as pertains to this "quality of driving" factor is the link between age and driving behavior. In the chapters that follow, we convey various benchmarking methods, including those that take into consideration driving quality. We also provide guidance to analysts on how to consider a potential "intervening variable": age.

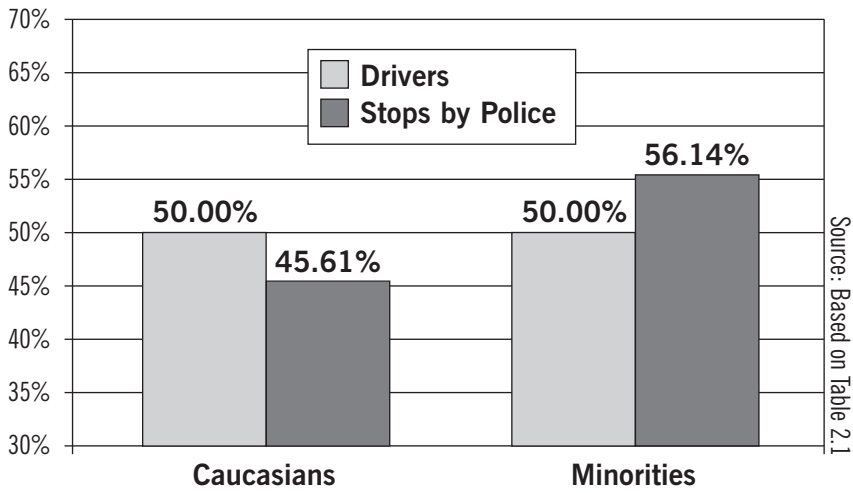


Figure 2.3. False Indication of Racial/Ethnic Bias Based on Age Differences of Drivers in Hypothetical Jurisdiction Q

Hypothesis 4: Racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.

The law enforcement activities of police are not the same in all areas at all times. Indeed, the level of vehicle stops by police may vary quite legitimately from area to area.[11] People who drive in areas where stopping activity by police is high are at greater risk of being stopped than are their counterparts who drive in areas with low stopping activity. This difference could affect efforts to assess racially biased policing if law enforcement activities vary across geographic areas where the demographic composition also varies. If variations in police stopping activity are not considered in analyses of police-citizen contact data, results that indicate disparity may reflect not racial/ethnic bias, but very legitimate variations in police practices.[12]

A hypothetical example, analogous to the earlier example that focused on differences in age demographics across racial/ethnic groups, illustrates how misleading indicators of racial/ethnic disparity can easily emerge. This example also highlights the need for researchers to conduct analyses within subareas of the jurisdiction under study. Table 2.2 shows the racial/ethnic profile of driving-age residents and the racial/ethnic profile of the drivers stopped in hypothetical Jurisdiction R (composed of Area A and Area B). There are an equal number of people of driving age in each area (1,000 each), but Area A is predominantly Caucasian (80 percent of driving-age residents)

---

11 Heavy levels of police deployment will not necessarily coincide with high levels of vehicle stops for traffic violations. In fact, in some high-crime areas where police deployment is likely to be correspondingly high, traffic enforcement may be a low priority in light of the more critical problems that need to be addressed.

12 These variations in police activities across areas within a jurisdiction would not be legitimate if the differential enforcement were based on inappropriate factors such as racial/ethnic bias. To discern whether bias is a factor, the researcher could assess whether legitimate factors (such as calls for service, traffic accidents) adequately predict levels of law enforcement activities.

and Area B is predominantly composed of minorities (80 percent of driving-age residents). In each area, the demographic profile of the drivers stopped by police matches the demographic profile of the driving-age adults in the area. That is, in Area A, 80 percent of the residents are Caucasians, and 20 percent are minorities; similarly, 80 percent of the drivers stopped by police are Caucasians and 20 percent are minorities. (We use this particular benchmark, residential population, for purposes of making our point—not to promote it as a method.)

Table 2.2. Representation of Caucasian and Minority Drivers in the Driving Population and Population of Stopped Drivers, by Subarea, Hypothetical Jurisdiction R

Types of Drivers	Area A			
	No. of Driving-Age Residents	Percent of Residents	No. of Stops	Percent of Stops
Caucasians	800	80%	80	80%
Minorities	200	20%	20	20%
Total	1,000	100%	100	100%

Types of Drivers	Area B			
	No. of Driving-Age Residents	Percent of Residents	No. of Stops	Percent of Stops
Caucasians	200	20%	40	20%
Minorities	800	80%	160	80%
Total	1,000	100%	200	100%

Types of Drivers	Total Jurisdiction			
	No. of Driving-Age Residents	Percent of Residents	No. of Stops	Percent of Stops
Caucasians	1,000	50%	120	40%
Minorities	1,000	50%	180	60%
Total	2,000	100%	300	100%

In Area B, like Area A, the demographic profile of the drivers stopped by police matches the demographic profile of the residents. In short, the results as analyzed within Area A and Area B indicate no disparity. Note, however, that more traffic stops are made in Area B than in Area A. The reason for the greater traffic enforcement within Area B in hypothetical Jurisdiction R is the occurrence of many accidents there, prompting concerned citizens to request that local law enforcement crack down on speeders. Because of this heightened traffic enforcement—legitimate in our example—twice as many stops are made in Area B (200 stops) than in Area A (100 stops). If the researcher had not controlled for police activity within the two areas but instead had presented data for the whole jurisdiction, a false disparity would have become evident. The researcher would have reported disproportionate representation of minorities among drivers stopped by police (see the Total Jurisdiction results of Table 2.2). When the absolute numbers of stops across areas are summed, and the demographic profile of the drivers who are stopped is compared to the demographic profile of the residential population, these misleading indications of disparity emerge. Those misleading data, graphed in Figure 2.4, show that minorities comprise 50 percent of the jurisdiction population but 60 percent of all stops. These would be the misleading results *even if* officers' decisions to stop were devoid of bias, and the increased traffic enforcement activity in Area B was completely legitimate.

In sum, it is appropriate to assume that people who drive in areas where stopping activity by police is high are at greater risk of being stopped than those who drive in areas where stopping activity is low. The nature and extent of policing activities may legitimately vary across geographic areas where the demographic composition also varies. Because of these possibilities, the methods used to analyze police-citizen contact data should reflect consideration of the hypothesis that *racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.* Because law enforcement agencies

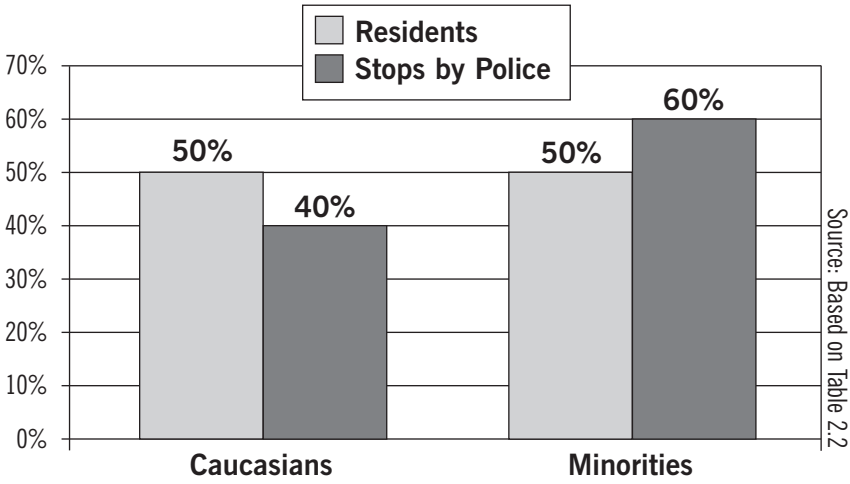


Figure 2.4. False Indication of Racial/Ethnic Bias Based on Differential Stopping Activity by Police Across Subareas in Hypothetical Jurisdiction R

cannot feel confident about the null hypothesis (there are no differences), they should take into account differential stopping activity by police across geographic areas when they analyze police-citizen contact data. In the chapters that follow, we discuss how researchers can recognize this alternative hypothesis. The example given here supports our recommendation that law enforcement agencies conduct analyses within geographic subareas of their jurisdiction and that they select those subareas in a way that allows researchers to hold constant (or "control for") the exposure of drivers to stopping activity by police.

SUMMARY OF THE BENCHMARKING CHALLENGE

The researcher developing a "benchmark" for police-citizen contact data is trying to determine the demographics (particularly the racial/ethnic composition) of the drivers who are at risk of being stopped, assuming no bias by police. We identified the key factors that influence this risk: driving quantity, driving quality, and the location of driving vis-a-vis levels of stopping activity by police. In order to determine whether there is a

cause-and-effect relationship between the race/ethnicity of drivers and police stopping behavior, we must be able to show that this relationship exists even when the other factors are considered. To test the hypothesis that driver race/ethnicity has an impact on stopping behavior by police, the alternative hypotheses that reflect the factors that increase the risk of being stopped must be ruled out. The alternative hypotheses are

- racial/ethnic groups are not equally represented as residents in the jurisdiction,
- racial/ethnic groups are not equally represented as drivers on jurisdiction roads,
- racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior, and
- racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.

It is not difficult to measure whether there is disparity between racial/ethnic groups in terms of stops made by police; the difficulty comes in identifying the causes for disparity. The alternative hypotheses present potential causes that need to be ruled out before a researcher can claim that the identified disparity is likely the result of police bias. After controlling for driving quantity, driving quality, and driving location (as pertains to levels of police stopping activity), a researcher who finds that minorities are disproportionately represented among drivers stopped by police can conclude with reasonable confidence that the disparity reflects police bias in their decision making. If no disparity was found, the researcher can fairly confidently conclude that bias was not a part of police decision making. If, on the other hand, the researcher finds disparity in the results after controlling for only driving quantity and driving location, he or she can report that disparity exists and that the results can be explained either by police bias or differential driving quality. That is, the researcher could not pinpoint a single cause (for example, bias) but must report that two possible explanations for the disparity remain.

Even results showing no disparity would need to be qualified if all factors were not controlled for. If, for instance, results indicated no disparity in stops, but driving quality had not been considered, the researcher cannot rule out the possibility of racial/ethnic bias in stopping behavior. We explore this possibility further in our discussion below of "masking."

A benchmark's value depends on the extent to which it addresses the alternative hypotheses. The higher the quality of the benchmark, the more confidence a researcher can have in the results. The need to rule out alternative hypotheses shows how much more complex benchmarking is than many have previously thought. When researchers attempt to interpret police-citizen contact data, they are, in effect, trying to look inside the heads of officers to discern their decision-making processes. Even a research model that incorporates the factors above does not begin to do justice to the complexity of these decisions. This caveat, however, is not unique to the analysis or interpretation of police-citizen contact data but is applicable to virtually all efforts by social scientists to measure human behavior and interaction.

THE PROBLEM OF INCONCLUSIVE RESULTS: A CENSUS BENCHMARKING EXAMPLE

In this section we use the census benchmarking method of analyzing police-citizen contact data to illustrate how researchers' failure to address the alternative hypotheses can lead to inconclusive results. In census benchmarking, a jurisdiction compares the demographic profile of the drivers stopped by police to the demographic profile of the residents of the jurisdiction as measured by the U.S. Decennial Census. Regardless of the results of this comparison (minorities are over-represented, minorities are under-represented, minorities are proportionately represented), *researchers can draw no definitive conclusions regarding racially biased policing.*

As an example, suppose that a law enforcement agency finds that minorities are over-represented among drivers

stopped by police relative to minorities' representation among jurisdiction residents. The racial/ethnic disparities manifested in this comparison might reflect racially biased policing, or they might reflect variation in the demographic profiles of (1) drivers on jurisdiction roads, (2) traffic law violators, or (3) drivers driving in locations where stopping activity by police is high. Our comparison of stop data to census data has indicated disparity, but the *causes* of that disparity have not been identified. We know that we have "disparate impact" (using the social science rather than the legal definition of the phrase), but we do not know if we have unjustified disparate impact in the form of racially biased policing. Because of these limitations, no conclusions can be drawn with regard to the existence or absence of racially biased policing.

Census benchmarking (assuming no adjustments of the census data)[13] takes into consideration only one of the four alternative hypotheses presented in this chapter—the hypothesis that *racial/ethnic groups are not equally represented as residents in the jurisdiction*.[14] Census benchmarking does not address hypotheses related to demographic variations across driving quantity, quality, or location. Nevertheless, stakeholders (for example, public officials, law enforcement executives, civil rights group representatives) often draw inappropriate conclusions about the results. Some of those inappropriate conclusions are represented in the benchmarking "myths" to which we turn next.

---

13 Chapter 5 discusses ways that census data are being adjusted by researchers in an attempt to encompass factors related to several, additional alternative hypotheses.

14 A common criticism of census data is the systematic undercounting of certain racial and ethnic groups. For the 2000 Census, the Census Bureau estimates that for one minority group—non-Hispanic Blacks—the percent undercount is statistically different from zero. The Bureau estimates a 1.84 percent undercount. See www.census.gov/dmd/www/ace2.html.

BENCHMARKING MYTHS

Myth 1: No racial/ethnic disparity means no racially biased policing.

As noted in the preceding example, the results produced by unadjusted census benchmarking, regardless of whether they showed under-representation, over-representation, or proportionate representation of minorities among the persons stopped by police, cannot enable researchers to draw sound conclusions about racially biased policing. This important truth has been contradicted in a few reports. Although the authors of these reports correctly acknowledge that their benchmarking method (census benchmarking) cannot produce conclusions regarding *the existence* of racially biased policing (because the alternative hypotheses have not been ruled out), they argue that it can prove *the absence* of racially biased policing. A finding of disproportionately high minority representation among persons stopped does not prove racially biased policing, they say, but a finding of disproportionately low minority representation or proportionate minority representation does prove that racially biased policing does not exist. This argument—that a method is valid for one result although not for another—is not true.

The adequacy of a law enforcement agency's benchmark is the same for all results. The researchers who put forth the argument that, regardless of benchmark quality, a showing of no disparity means no racially biased policing fail to recognize that an inadequate benchmark can "mask" (or hide) disparity. The following example shows how.

Let us say that a jurisdiction uses census benchmarking and finds that the demographic profile of residents matches perfectly the demographic profile of people stopped by police. It is still possible that policing in the jurisdiction is racially biased. If minorities are on the road in, or violating at, proportions less than their residential representation, the fact that they are stopped proportionate to their residential representation indicates disparity, and it may indicate racially biased policing. Indeed, the existence of racially biased policing may be masked

by flaws inherent in the benchmark. Hypothetical data on the representation of minorities and nonminorities among jurisdiction residents, traffic violators, and people stopped for traffic violations are presented in Figure 2.5. It shows that 25 percent of the residents are racial/ethnic minorities as are 25 percent of the people who are stopped by police for traffic violations. This is the type of finding (a finding of no disparity) that some mistakenly have argued indicates an absence of racially biased policing.

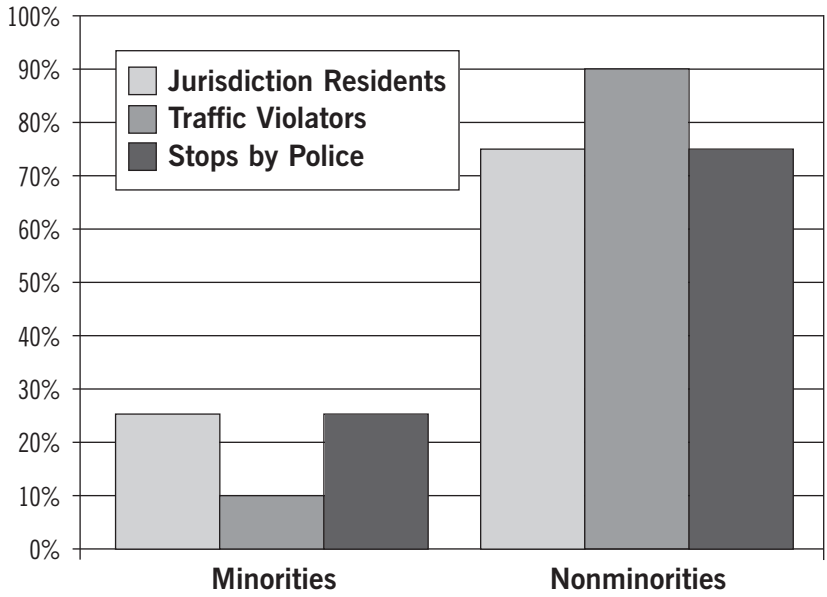


Figure 2.5. Racially Biased Policing Masked in Hypothetical Jurisdiction S

The figure also shows the proportion of minorities and nonminorities who are *traffic violators* (information that would not be available to the researcher who conducted only census benchmarking), and this information indicates that minorities are over-represented among the drivers who are stopped. If minorities comprise only 10 percent of the traffic violators (that

is, 10 percent of the population *legitimately at risk of being stopped by police*), but 25 percent of the population that *is stopped by police*, racial bias is indicated. The key here is that the researcher conducting census benchmarking would not have had the information (on violating behavior) necessary to interpret either results that showed disparity or results that showed no disparity.

Researchers who are assessing police-citizen contact data should remember that (1) a weak benchmark is weak for all results, and (2) their benchmarking method can mask racially biased policing.

Myth 2: Results from a weak methodology become more worthy over time.

It is not true that results from a weak methodology, or benchmark, can become a worthy baseline for interpreting data in subsequent years—at least not for the purpose of assessing the existence of racially biased policing. An example will help explicate this myth. Let's say that a jurisdiction uses census benchmarking and determines that racial/ethnic minorities are over-represented among people stopped by police relative to their representation in the residential population as measured by the census. As explained above, these results indicate the existence of a disparity but not its cause. The temptation for stakeholders, and even some researchers, is to equate the disparity with racially biased policing and to desire a reduction in that disparity in subsequent years. That is, they might acknowledge that their benchmark is weak, but claim nonetheless that the results produced during the first year of analysis can be used to assess and evaluate change in subsequent years. This is not true. Because of the weak methods used, the researcher cannot equate the disparity with racially biased policing and therefore should not presume that a reduction in disparity the following year would be desirable and that it would indicate reduced bias. The disparity may reflect wholly legitimate factors at work. If that is the case (which cannot be known with

some benchmarking methods), then a reduction in disparity is not a legitimate goal.

Similarly, a jurisdiction that finds no disparity as a result of its census benchmarking analysis the first year and does find disparity the second year should not blame the police department. Again, because of the methods used, this disparity cannot be equated with police bias. In sum, a benchmark that cannot pinpoint cause cannot produce explanations of cause over time.

Myth 3: Results from a weak methodology become strong if replicated in multiple geographic areas.

A police department that conducts census benchmarking within multiple subareas of the city (say, within each police district) and finds no evidence of racial/ethnic disparity in each one can easily believe the myth stated above. The police spokesperson might acknowledge the weaknesses of census benchmarking but discount those weaknesses and claim that because the results are consistent throughout the city, this proves policing in the city is not racially biased. Such a claim would be in error. The results from a weak methodology are not validated if the results are consistent across multiple geographic areas.

If a methodology can measure only disparity and not the cause of that disparity, that limitation persists even when the methodology is used over and over again in multiple areas. In a contrasting example, a researcher may find disparity in all or most of the subareas within a jurisdiction. Again, however, multiple measures of disparity do not accumulate to provide a cause for that disparity; they continue to represent only multiple measures of disparity.

CONCLUSION

In this chapter we discussed the challenge of benchmarking—the process of developing a demographic profile of drivers at risk of being stopped by police, assuming no bias. We pinpointed the factors that should legitimately increase or decrease the likelihood of being stopped and framed those factors in the form

of "alternative hypotheses" to the "bias hypothesis." To assess whether there is differential stopping by police of demographic groups, we test the hypothesis that police are biased in their decision making and do so by ruling out the alternative hypotheses. The strength of a benchmark depends on the degree to which it encompasses the factors associated with the alternative hypotheses. In Chapters 5 through 10 we discuss the major benchmarking methods: adjusted census benchmarking, benchmarking based on a comparison of licensed drivers and drivers stopped by police, benchmarking based on blind versus not-blind enforcement, internal benchmarking, and observation-based benchmarking. The framework of alternative hypotheses is used to convey the strength of the benchmark and, relatedly, to make recommendations regarding how the results of the police-citizen contact data analysis can be responsibly conveyed. However, before we turn to these various benchmarking methods, we discuss how agencies mandated or choosing to collect data initiate collection (Chapter 3) and prepare the data for analysis (Chapter 4).

II Getting Started

This chapter describes the preliminary steps associated with collecting police-citizen contact data and explains how and why a jurisdiction might involve residents, police personnel from all levels of the department, and independent social scientists in these efforts. Additionally, we discuss factors that a law enforcement agency should consider before choosing a benchmark for analyzing its data.

Any law enforcement agency that is planning to collect data needs to address the following questions:

- On what law enforcement activities should the agency collect data?
- What information should the agency collect regarding those activities?
- How should the agency analyze and interpret the data?

Building upon the work of Ramirez, McDevitt, and Farrell (2000), Fridell et al. (2001, Chap. 8) discuss the options available to agencies regarding the first two questions. For instance, the 2001 report reviews the considerations for deciding whether to collect data on traffic stops only, all vehicle stops, or all detentions (including pedestrian stops). Also discussed are the data elements that agencies should consider for inclusion in

their protocol (for example, the date, time, and reason for the vehicle stop; the race, ethnicity, age, and gender of the person stopped; information regarding stop dispositions and search activity). We do not repeat those discussions here. Agencies in the first stages of planning data collection will find these previously published sources helpful. (Again, the Fridell 2001 document can be downloaded from www.policeforum.org.) It also may be constructive for them to contact peer agencies and request to review their "forms."[1] Be sure to ask relevant personnel what, in hindsight, they would change about their forms.

DEVELOPING THE DATA COLLECTION PROTOCOL: TWO RECOMMENDATIONS

We offer two important recommendations related to developing the data collection protocol. First, plans for how an agency will analyze its data should be developed, if feasible, at the same time the decision makers develop the overall data collection strategy. Uninformed or after-the-fact decisions in these matters can lead to unnecessary tensions between residents (particularly racial/ethnic minority residents) and policy makers and/or between police officers and policy makers. Both jurisdiction residents and officers have a strong stake in the highest quality analyses of the data. Officers, in particular, can be legitimately skeptical of—even strongly opposed to—data collection efforts if they lack assurances that the data will be analyzed using the best social science methods available or, at least, responsibly interpreted. An early designation of the method of analysis and a commitment to responsible interpretation can mitigate these concerns. In the same vein, it is important for the agency to confirm early on that sufficient resources are available to meet

---

1 Not all agencies are using paper forms to collect their data. Some agencies ask their officers to submit data by using handheld or in-car computers; in other agencies, officers verbally submit the stop information over the radio. The word "forms" used throughout this report denotes all methods of data submission.

its objectives. Otherwise, an agency may make a significant investment in a data collection system only to find out that analyses of the quality it desires cannot be implemented. Some of the methods that can be chosen to analyze police-citizen contact data rely on particular data elements in the forms that officers complete. This is another reason for early planning.

Second, we strongly advise that, in identifying which activities a jurisdiction will target for data collection, the decision makers select all traffic stops, all vehicle stops, and/or all stops including pedestrian stops and not a subset of any of these categories as defined by their outcomes.[2]

Some agencies (indeed, some states) are collecting and analyzing data only from the traffic stops that *result in citations*. (That is, instead of collecting and analyzing data from all traffic stops, these jurisdictions are focusing on a subset of traffic stops as defined by the outcome, a citation.) This common practice is convenient because it does not add paperwork for the officers (relying, as it does, on existing, albeit possibly modified, forms), but the practice is not recommended. The resulting data exclude stops by police that may be at heightened risk of being racially motivated. A data collection system based on citation stops alone excludes stops of law-violating drivers who should have received a citation but did not, and it may include law-abiding drivers who should not have been stopped in the first place. The "selection" by police of the fortunate drivers or illegitimately stopped drivers could be based on their race/ethnicity, and thus by excluding drivers who do not receive citations, a jurisdiction severely jeopardizes its ability to assess the existence of racially biased policing, regardless of the strength of the benchmark used. The

---

2 As explained in Chapter 1, "vehicle stop" denotes any stop made by police of a person in a vehicle; "traffic stop" denotes a vehicle stop the stated purpose of which is to respond to a violation of traffic laws (including codes related to quality/maintenance of vehicles); and "investigative (vehicle) stop" denotes police stops of people in vehicles when there is at least reasonable suspicion of criminal activity.

researcher could, with these limited data, identify bias where none exists or conclude there is no bias when, in fact, there is.

This faulty methodology is analogous to assessing the impact of race on prison sentences by focusing only on those who are in prison. For example, by examining only the racial makeup of the prison population and comparing length of prison sentences across races, a jurisdiction will be unable to reach sound conclusions. It must also assess whether or not there are racial differences with regard to who gets sentenced to prison (versus sentenced to jail or to probation, for example).

If a jurisdiction is collecting data only on subsets of stops, it needs to include a strongly stated caveat regarding the stops that are excluded from its research. This limitation on the data concerning who is stopped will also affect the analysis of post-stop activities and outcomes. This is because some people who were stopped by police—some of whom were searched and maybe even detained for long periods of time—will not be included in the data set being analyzed.

INVOLVING RESIDENTS AND POLICE PERSONNEL IN PLANNING DATA COLLECTION AND ANALYSIS

It is advantageous for jurisdictions to involve residents and a cross-section of law enforcement agency employees in planning how the data will be collected and analyzed. (Regarding the latter, we note that even if a jurisdiction did not involve residents and police in planning the data collection system, it could still involve them in discussions about the data's analysis and interpretation.)

Police personnel—particularly line personnel—can bring valuable information and an important perspective to the table. These agency representatives have a critical stake in ensuring a high-quality initiative, and they should have the opportunity to raise any of their concerns about the integrity and fairness of the data collection and analysis system. Employees' involvement can also facilitate "buy in" by the line officers upon whom the agency will rely to collect the data.

Law enforcement agencies' involvement of residents (particularly minority residents) in data collection planning can improve police-citizen relations, enhance the credibility of the research efforts, and increase the likelihood that the community will view the outcome as legitimate.[3] Involving jurisdiction residents in discussions regarding data analysis/interpretation has the additional advantage of educating a core group of residents about the complexities and constraints of the process. These residents can serve as important voices affirming the integrity of the analysis and the sound interpretation of the results when reports are released to the public.

In the interest of responsible social science, the caveats associated with various benchmarking methods should be included in jurisdiction reports. The caveats should convey why the results may not provide definitive proof of racially biased policing or its absence in the jurisdiction. Coming only from the police department spokesperson, these caveats may be interpreted by skeptical residents as defensive excuses for why results showing disparity (if they do) are not proof of racial bias. Although the use of independent social scientists to conduct analyses will add credibility to these caveats, the additional voices of respected residents who understand the methodological constraints will increase the likelihood that the results and the conclusions drawn from them will be viewed as legitimate by the general public and the media. "If the community understands benchmarks and the variables that skew aggregate data there is less likelihood the information will be misinterpreted and misused," writes McMahon et al. (2002, 94). One way to facilitate the understanding of data analyses on the part of citizens is to set up a local racial profiling task force or advisory committee.

---

[3] See Farrell, McDevitt, and Buerger (2002) for a discussion of how police-community task forces can be used to oversee the data collection system and to otherwise address the issue of racially biased policing in a jurisdiction.

As recommended in PERF's first report on the topic of racially biased policing, these task forces should be composed of fifteen to twenty-five people with representatives from both the department and the community (Fridell et al. 2001, Chap. 7). In selecting community members, decision makers should focus on those people who are most concerned about racial bias by police. The task force should include representatives from the jurisdiction's various minority groups and representatives from civil rights groups. Consideration should be given to media representatives as well because these professionals will be in the important position of conveying the results to jurisdiction residents. Police personnel selected for the task force should represent all departmental levels, particularly patrol.

Citizens and police can bring knowledge to the discussions that is of value in planning the data analyses and understanding the results. What they know about the jurisdiction's characteristics, residents, and police activities can be of great help to the researchers charged with actually implementing the analysis plan. For instance, their knowledge of jurisdiction roads may be helpful to a researcher trying to choose representative intersections where observers will document the race/ethnicity of drivers. (See discussion of the observation method of benchmarking in Chapter 9.) Or their knowledge that a particular high-minority downtown entertainment area draws large numbers of white suburbanites on Saturday nights can be helpful to a researcher seeking to understand the results for that area.

PARTNERING WITH SOCIAL SCIENTISTS

If resources allow, an agency should consider obtaining the assistance of independent social scientists for analyzing its police-citizen contact data. There are two major reasons for partnering with social scientists:

- Partnering with an individual or a team external to the agency can add credibility to the process and thus to the results.

- The skills of trained social scientists can supplement the internal resources available for research.

Data collection to assess racially biased policing is both a social science and a political endeavor. Thus, an agency must attend to both social science and political objectives in developing and implementing an analysis plan. An agency could use internal staff to conduct a high-quality analysis but lose in the political arena because the jurisdiction's residents did not consider the internally conducted analysis to be credible.

Many law enforcement agencies (especially small and medium-sized ones) do not have the in-house expertise to analyze and interpret police-citizen contact data. A social science partner may be essential to supplement agency resources and perform these functions. The analyst(s) should be trained in social science methods and have general knowledge of law enforcement; they also should have demonstrated knowledge of the specific issues associated with analyzing police-citizen contact data (Fridell et al. 2001, Chap. 8). Ideally, this "demonstrated knowledge" would come from having conducted similar analyses for other jurisdictions. Capable analysts are most likely to be associated with a college or university or with an independent research firm. The individual social scientist or the research team will play a major role in educating jurisdiction residents about the various methods that can be used for analysis and the strengths and weaknesses of each.

Importantly, the social scientist(s) become "partners" with the agency or, preferably, with the jurisdiction task force in the data collection/analysis effort. They are not just handed the data to analyze as they see fit in the privacy of their university or agency offices. The analysis plan should be agreed upon by all parties and the social scientists should communicate with their agency and/or task force partners throughout their work. The researchers should share preliminary results, soliciting perspectives from their police and resident partners who likely have superior knowledge regarding local conditions that may be pertinent to the interpretation of the data.

SELECTING BENCHMARKS

In subsequent chapters we describe the various benchmarks that law enforcement agencies can use to analyze and interpret vehicle stop data. These benchmarks vary considerably in terms of their ability to address the alternative hypotheses discussed in Chapter 2. In deciding which benchmark(s) to use, decision makers should consider the following factors: the level of measurement precision they desire, the financial and personnel resources that are available, the data elements that must be collected, and the availability of other data that may be required for using a particular benchmark. Later chapters describe for each benchmarking method its level of precision, required agency resources, required data elements, and requirements in terms of information from outside sources.

Level of Measurement Precision Desired

The higher the quality of the benchmark, the greater the ability of the researcher to "measure" and draw conclusions regarding racially biased policing. High-quality analysis can provide meaningful information not only on whether the problem exists and, if so, to what degree, but also on the nature of the problem and the specifics of its manifestation (in terms of particular geographic areas, shifts, or officers). However, the institution conducting the analysis need not pick one of the most precise methodologies (coming as these do with generally higher complications and sometimes higher costs) in order to make its data collection system successful and constructive. The keys to success for an agency picking a benchmark are (1) responsible interpretation and (2) constructive discussion with stakeholders concerning benchmark weaknesses.

For each benchmark described in later chapters, we provide information related to the strength of the conclusions being drawn. (This will be conveyed in terms of the extent to which each benchmark encompasses the alternative hypotheses.) Reports will need to include this information to ensure responsible interpretation of the data. Imperfect data can still provide a

solid base for constructive dialogue between police and citizens. Results showing "disparity" that cannot be linked to a particular "cause" (such as bias) can still lead to a meaningful discussion of possible causes and desirable reforms. Importantly, these discussions can lead to the collection of other forms of "data," including that which comes from an open and frank sharing of concerns by citizens.[4] Commenting on the value of police-citizen contact data for facilitating police-citizen dialogue, Farrell, McDevitt, and Buerger (2002, 365) report: "The most effective and productive use of racial profiling data is not its ability to determine if racial profiling exists but rather its ability to provide concrete information to ground police-community discussions about patterns of stops, searches, and arrests throughout local communities."

Required Agency Resources

In selecting a benchmark for analyzing police-citizen contact data, an agency should consider not only the level of measurement precision it desires but also the resources it has available. Not surprisingly, the most effective benchmarks usually (but not necessarily) require the most resources in terms of finances and personnel. An agency will want to select the most effective method given its resources and objectives.[5]

Data Elements

The use of some benchmarks is dependent on the inclusion of particular elements on the data collection form. If the agency is

---

4 Fridell et al. (2001, Chap. 7) promotes police-resident discussions of racially biased policing and perceptions of its practice. A video and accompanying guide, funded by the Office of Community Oriented Policing Services in the U.S. Department of Justice, was developed to facilitate and structure these dialogues. This video and guide can be ordered through the PERF web site, www.policeforum.org.

5 We do not have reliable information regarding the costs that are associated with the various benchmarks. Many jurisdictions seeking to hire outside analysts issue requests for proposals and then review the proposals, balancing strength of methodology and resources required.

in the early stages of developing the data collection protocol, decisions regarding how to analyze/interpret the data should be made in conjunction with decisions about the content of the form (that is, what data elements to include). If an agency has already developed the form, decision makers will need to ensure that the method selected for analysis/interpretation is supported by available data. As an example, we describe in Chapter 7 how some jurisdictions have compared the demographic profiles of drivers stopped for speeding by police unaided by radar to the demographic profiles of drivers stopped because of radar measurements of their speed. (The radar stops are conducted in a manner so that the radar operator cannot discern the driver's race/ethnicity.) To make such a comparison, the jurisdiction must be able to identify, from data on the forms, which stops were conducted with and without radar.

For all benchmarking methods we advocate analyses within specific geographic subareas. Therefore, the location of the stop is an important data element to include on the police-citizen contact data form. For purposes of reviewing and monitoring data for quality, a unique identifier (number) on the form also is helpful. Most advantageous is an incident number or similar identifier that corresponds to information about the event that is contained in other data sets, such as computer-aided-dispatch (CAD) data and citation data.

The Availability of Other Data

Some benchmarking methods are dependent upon the availability of information from outside sources. An example is a method that compares the demographic profile of drivers who are identified as traffic violators by enforcement cameras (cameras that are used at controlled jurisdictions to detect and ticket red-light violators or speeders) to the demographic profile of drivers who are identified as traffic violators by officers on patrol in the same area as the cameras. This method would, of course, be available only to jurisdictions that have enforcement cameras in place and are able to identify through the license

plate number (or photos) the race/ethnicity of the violators (or at least the race/ethnicity of the vehicle owners when the license plate number is used).

Other Considerations

A jurisdiction may decide to use multiple benchmarks. For example, it might implement "internal" benchmarking and some "external" method as well. Internal benchmarking is a strong benchmark for identifying which police officers, units, or shifts may be stopping minorities at higher rates than their "similarly situated" counterpart officers, units, or shifts. A drawback to internal benchmarking, however, is that it only compares parts of the law enforcement agency to itself. For this reason, the agency might choose—in addition—to compare the agency's performance to some outside benchmark, such as that provided by the blind versus not-blind enforcement method, or the observation method. Thus, a jurisdiction might implement both internal benchmarking and some external method as well.

An agency might also decide to implement a relatively simple benchmark (for example, adjusted census benchmarking) in all the subareas of its jurisdiction and then invest in a more complicated and more effective benchmark (for example, the observation methodology described in Chapter 9) in those subareas identified by the simpler benchmark as having the greatest racial/ethnic disparities.

INFORMING THE PUBLIC OF DATA COLLECTION EFFORTS

Some law enforcement executives, when announcing their data collection efforts, have referred to the initiative as an opportunity to "prove" that policing in their jurisdiction is not racially biased. This is inappropriately and unnecessarily defensive. First of all, such a prediction of research results is *inappropriate*. While a particular executive might be justified in having confidence that racially biased policing is neither systematic nor widespread within his or her jurisdiction, the executive is

naïve to claim absolutely that it never occurs. Such a statement is almost certain to offend racial/ethnic minorities who perceive otherwise. Our society has serious racial/ethnic biases, and the police profession—like every other profession—hires from a population with these prejudices. Even in a department in which racial bias is neither systematic nor widespread, it is likely that it occurs in some places, at some times, committed by some individual officers. Finally, such a strong claim (the police executive's use of the word "prove") implies that police-citizen contact data can provide definitive answers—which they cannot. As is true of social science in general, even strong methods will not provide definitive proof of the existence or lack of racially biased policing.

A claim of innocence even before the data are collected and analyzed is also *unnecessary*. An executive can reasonably assert that the agency is undertaking data collection in a sincere effort to determine whether or not a problem of racial/ethnic bias exists and, if it does, will implement corrective and preventive actions.

That said, we are not advocating that agencies wait until the data are collected and analyzed to implement remedial actions. In a perfect world (where social science could quickly and definitively answer all the questions we pose), agencies would first analyze the problem and then, based on that analysis, develop appropriate responses (policies, training, outreach) to promote reform. In the context of our imperfect world (where data collection takes time and social science cannot provide definitive results), agencies should not make data collection showing racial disparity the minimum requirement for implementing reforms to address this critical issue. In fact, while the practice of data collection as a response to racially biased policing has had important benefits, a negative side effect, arguably, is the inherent implications that (1) some agencies are "guilty" of racial bias and others are not and (2) agencies shown to be "guilty" are the ones that should implement reforms. All agencies committed to democratic policing,

not only agencies "proven guilty" of bias through data collection, need to implement reforms.[6]

CONCLUSION

In this chapter we reviewed important considerations in developing the data form and deciding which types of activities to target for data collection. We encourage the involvement of residents and police personnel from all levels in making decisions regarding the data collection system, and we discussed the circumstances in which agencies might want to involve independent social scientists. The selection of benchmarks should be based on considerations of measurement precision, resources, existing data elements, and the availability of other data. A police executive announcing data collection plans to the public should not claim innocence before the fact. Indeed, like society at large, an agency is rarely bias free. Neither should that agency executive await the results of data collection—whatever they might be—to implement reforms to address the long-standing, widespread issues of racially biased policing and the perceptions of its practice.

---

6 Various responses to racially biased policing are set forth in Fridell et al. (2001).